Fourier Analysis of Stationary Processes

DAVID R. BRILLINGER

Abstract—This paper begins with a description of some of the important procedures of the Fourier analysis of real-valued stationary discrete time series. These procedures include the estimation of the power spectrum, the fitting of finite parameter models, and the identification of linear time invariant systems. Among the results emphasized is the one that the large sample statistical properties of the Fourier transform are simpler than those of the series itself. The procedures are next generalized to apply to the cases of vector-valued series, multidimensional time series or spatial series, point processes, random measures, and finally to stationary random Schwartz distributions. It is seen that the relevant Fourier transforms are evaluated by different formulas in these further cases, but that the same constructions are carried out after their evaluation and the same statistical results hold. Such generalizations are of interest because of current work in the fields of picture processing and pulse-code modulation.

I. INTRODUCTION

THE FOURIER analysis of data has a long history, dating back to Stokes [1] and Schuster [2], for example. It has been done by means of arithmetical formulas (Whittaker and Robinson [3], Cooley and Tukey [4]), by means of a mechanical device (Michelson [5]), and by means of real-time filters (Newton [6], Pupin [7]). It has been carried out on discrete data, such as monthly rainfall in the Ohio valley (Moore [8]), on continuous data, such as radiated light (Michelson [5]), on vector-valued data, such as vertical and horizontal components of wind speed (Panofsky and McCormick [9]), on spatial data, such as satellite photographs (Leese and Epstein [10]), on point processes, such as the times at which vehicles pass a position on a road (Bartlett [11]), and on

point processes in space, such as the positions of pine trees in a field (Bartlett [12]). It has even been carried out on the logarithm of a Fourier transform (Oppenheim *et al.* [13]) and on the logarithm of a power spectrum estimate (Bogert *et al.* [14]).

The summary statistic examined has been: the Fourier transform itself (Stokes [11]), the modulus of the transform (Schuster [2]), the smoothed modulus squared (Bartlett [15]), the smoothed product of two transforms (Jones [16]), and the smoothed product of three transforms (Hasselman *et al.* [17]).

The summary statistics are evaluated in an attempt to measure population parameters of interest. Foremost among these parameters is the power spectrum. This parameter was initially defined for real-valued-time phenomena (Wiener [18]). In recent years it has been defined and shown useful for spatial series, point processes, and random measures as well. Our development in this paper is such that the definitions set down and mathematics employed are virtually the same for all of these cases.

Our method of approach to the topic is to present first an extensive discussion of the Fourier analysis of real-valued discrete-time series emphasizing those aspects that extend directly to the cases of vector-valued series, of continuous spatial series, of point processes, and finally of random distributions. We then present extensions to the processes just indicated. Throughout, we indicate aspects of the analysis that are peculiar to the particular process under consideration. We also mention higher order spectra and nonlinear systems. Wold [19] provides a bibliography of papers on time series analysis written prior to 1960. Brillinger [20] presents a detailed description of the Fourier analysis of vector-valued discrete-time series.

We now indicate several reasons that suggest why Fourier analysis has proved so useful in the analysis of time series.

II. Why the Fourier Transform?

Several arguments can be advanced as to why the Fourier transform has proved so useful in the analysis of empirical functions. For one thing, many experiments of interest have the property that their essential character is not changed by moderate translations in time or space. Random functions produced by such experiments are called *stationary*. (A definition of this term is given later.) Let us begin by looking for a class of functions that behave simply under translation. If, for example, we wish

$$f(t + u) = C_u f(t), \qquad t, u = 0, \pm 1, \pm 2, \cdots$$

with $C_1 \neq 0$, then by recursion

$$f(t) = C_1 f(t - 1) = C_2 f(t - 2) = \cdots = C_1^t f(0)$$

for $t \geq 0$ and so $f(t) = f(0) \exp \{\alpha t\}$ for $\alpha = \ln C_1$. If $f(t)$ is to be bounded, then $\alpha = i\lambda$, for $i = \sqrt{-1}$ and λ real. We have been led to the functions $\exp \{i\lambda t\}$. Fourier analysis is concerned with such functions and their linear combinations.

On the other hand, we might note that many of the operations we would like to apply to empirical functions are linear and translation invariant, that is such that; if $X_1(t) \to Y_1(t)$ and $X_2(t) \to Y_2(t)$ then $\alpha_1 X_1(t) + \alpha_2 X_2(t) \to \alpha_1 Y_1(t) + \alpha_2 Y_2(t)$ and if $X(t) \to Y(t)$ then $X(t - u) \to Y(t - u)$. Such operations are called *linear filters*. It follows from these conditions that if $X(t) = \exp \{i\lambda t\} \to Y_\lambda(t)$ then

$$X(t + u) = \exp \{i\lambda u\} X(t) \to \exp \{i\lambda t\} Y_\lambda(t) = Y(t + u).$$

Setting $u = t$, $t = 0$ gives $Y_\lambda(t) = \exp \{i\lambda t\} Y_\lambda(0)$. In summary, $\exp \{i\lambda t\}$ the complex exponential of frequency λ is carried over into a simple multiple of itself by a linear filter. $A(\lambda) = Y_\lambda(0)$ is called the *transfer function* of the filter. If the function $X(t)$ is a Fourier transform, $X(t) = \int \exp \{i\alpha t\} x(\alpha) d\alpha$, then from the linearity (and some continuity) $X(t) \to \int \exp i\alpha t \, A(\alpha) x(\alpha) d\alpha$. We see that the effect of a linear filter is easily described for a function that is a Fourier transform.

In the following sections, we will see another reason for dealing with the Fourier transforms of empirical functions, namely, in the case that the functions are realizations of a stationary process, the large sample statistical properties of the transforms are simpler than the properties of the functions themselves.

Finally, we mention that with the discovery of fast Fourier transform algorithms (Cooley and Tukey [4]), the transforms may often be computed exceedingly rapidly.

III. Stationary Real-Valued Discrete-Time Series

Suppose that we are interested in analyzing T real-valued measurements made at the equispaced times $t = 0, \cdots, T - 1$. Suppose that we are prepared to model these measurements by the corresponding values of a realization of a stationary discrete-time series $X(t)$, $t = 0, \pm 1, \pm 2, \cdots$. Important parameters of such a series include its *mean*,

$$c_X = EX(t) \tag{1}$$

giving the average level about which the values of the series are distributed and its *autocovariance function*

$$c_{XX}(u) = \text{cov} \{X(t + u), X(t)\}$$
$$= E \{[X(t + u) - c_X] [X(t) - c_X]\}, \qquad u = 0, \pm 1, \cdots \tag{2}$$

providing a measure of the degree of dependence of values of the process $|u|$ time units apart. (These parameters do not depend on t because of the assumed stationarity of the series.) In many cases of interest the series is *mixing*, that is, such that values well separated in time are only weakly dependent in a formal statistical sense to be described later. Suppose, in particular, that $c_{XX}(u) \to 0$ sufficiently rapidly as $|u| \to \infty$ for

$$f_{XX}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} c_{XX}(u) \exp \{-i\lambda u\}, \quad -\infty < \lambda < \infty \tag{3}$$

to be defined. The parameter $f_{XX}(\lambda)$ is called the *power spectrum* of the series $X(t)$ at frequency λ. It is symmetric about 0 and has period 2π. The definition (3) may be inverted to obtain the representation

$$c_{XX}(u) = \int_{-\pi}^{\pi} \exp \{i\alpha u\} f_{XX}(\alpha) d\alpha \tag{4}$$

of the autocovariance function in terms of the power spectrum.

If the series $X(t)$ is passed through the linear filter

$$X(t) \to Y(t) = \sum_u a(t - u) X(u)$$

with well-defined transfer function

$$A(\lambda) = \sum_u a(u) \exp \{-i\lambda u\}$$

then we can check that

$$c_{YY}(u) = \sum_u \sum_v a(u + v) a(w) c_{XX}(w - v) \tag{5}$$

and, by taking Fourier transforms, that

$$f_{YY}(\lambda) = |A(\lambda)|^2 f_{XX}(\lambda) \tag{6}$$

under some regularity conditions. Expression (6), the frequency domain description of linear filtering, is seen to be much nicer than (5), the time-domain description.

Expressions (4) and (6) may be combined to obtain an interpretation of the power spectrum at frequency λ. Suppose that we consider a narrow band-pass filter at frequency λ having transfer function

$$A(\alpha) \doteq \begin{cases} 1, & |\alpha \pm \lambda| \leq \Delta \\ 0, & \text{otherwise} \end{cases}$$

with Δ small. Then the variance of the output series $Y(t)$, of the filter, is given by

$$\text{var } Y(t) = c_{YY}(0)$$
$$= \int f_{YY}(\alpha) d\alpha$$
$$= \int |A(\alpha)|^2 f_{XX}(\alpha) d\alpha$$
$$= 4\Delta f_{XX}(\lambda). \tag{7}$$

In words, the power spectrum of the series $X(t)$ at frequency λ is proportional to the variance of the output of a narrow band-pass filter of frequency λ. In the case that $\lambda \neq 0, \pm 2\pi, \pm 4\pi, \cdots$

196

the mean of the output series is 0 and the variance of the output series is the same as its mean-squared value. Expression (7) shows incidentally that the power spectrum is nonnegative.

We mention, in connection with the representation (4), that Khintchine [21] shows that for $X(t)$ a stationary discrete time series with finite second order moments, we necessarily have

$$c_{XX}(u) = \int_{-\pi}^{\pi} \exp\{i\alpha u\} \, dF_{XX}(\alpha) \qquad (8)$$

where $F_{XX}(\alpha)$ is a monotonic nondecreasing function. $F_{XX}(\lambda)$ is called the *spectral measure*. Its derivative is the power spectrum. Going along with (8), Cramér [22] demonstrated that the series itself has a Fourier representation

$$X(t) = \int_{-\pi}^{\pi} \exp\{i\alpha t\} \, dZ_X(\alpha), \qquad t = 0, \pm 1, \cdots \qquad (9)$$

where $Z_X(\lambda)$ is a random function with the properties;

$$E \, dZ_X(\lambda) = \eta(\lambda) \, c_X \, d\lambda \qquad (10)$$

$$\text{cov}\{dZ_X(\lambda), dZ_X(\mu)\} = \eta(\lambda - \mu) \, dF_{XX}(\lambda) \, d\mu. \qquad (11)$$

(In these last expressions, if $\delta(\lambda)$ is the Dirac delta function then $\eta(\lambda) = \Sigma \, \delta(\lambda - 2\pi j)$ is the Kronecker comb.) Also expression (11) concerns the covariance of two complex-varied variates. Such a covariance is defined by $\text{cov}\{X, Y\} = E\{(X - EX)\overline{(Y - EY)}\}$.) Expression (9) writes the series $X(t)$ as a Fourier transform. We can see that if the series $X(t)$ is passed through a linear filter with transfer function $A(\lambda)$, then the output series has Fourier representation

$$\int_{-\pi}^{\pi} \exp\{i\alpha t\} \, A(\alpha) \, dZ_X(\alpha), \qquad t = 0, \pm 1, \cdots.$$

In Section XV, we will see that the first and second-order relations (10), (11) may be extended to kth order relations with the definition of kth order spectra.

IV. THE FINITE FOURIER TRANSFORM

Let the values of the series $X(t)$ be available for $t = 0, 1, 2, \cdots, T - 1$ where T is an integer. The *finite Fourier transform* of this stretch of series is defined to be

$$d_X^{(T)}(\lambda) = \sum_{t=0}^{T-1} X(t) \exp\{-i\lambda t\}, \qquad -\infty < \lambda < \infty. \qquad (12)$$

A number of interpretations may be given for this variate. For example, suppose we take a linear filter with transfer function concentrated at the frequency λ, namely $A(\alpha) = \delta(\alpha - \lambda)$. The corresponding time domain coefficients of this filter are

$$a(u) = (2\pi)^{-1} \int A(\alpha) \exp\{iu\alpha\} \, d\alpha$$

$$= (2\pi)^{-1} \exp\{iu\lambda\}, \qquad u = 0, \pm 1, \cdots.$$

The output of this filter is the series

$$(2\pi)^{-1} \sum_u X(u) \exp\{i\lambda(t - u)\} \doteq (2\pi)^{-1} \exp\{i\lambda t\} \, d_X^{(T)}(\lambda).$$

These remarks show that the finite Fourier transform may be interpreted as, essentially, the result of narrow band-pass filtering the series.

Before presenting a second interpretation, we first remark that the sample covariance of pairs of values $X(t)$, $Y(t)$, $t = 0$, $1, \cdots, T - 1$ is given by $T^{-1} \Sigma X(t) Y(t)$, when the $Y(t)$ values have 0 mean. This quantity is a measure of the degree of linear relationship of the $X(t)$ and $Y(t)$ values. The finite Fourier transform is essentially, then, the sample covariance between the $X(t)$ values and the complex exponential of frequency λ. It provides some measure of the degree of linear relationship of the series $X(t)$ and phenomena of exact frequency λ.

In the case that $\lambda = 0$, the finite Fourier transform (12) is the sample sum. The central limit theorem indicates conditions under which a sum of random variables is asymptotically normal as the sample size grows to ∞. Likewise, there are theorems indicating that $d_X^{(T)}(\lambda)$ is asymptotically normal as $T \to \infty$. Before indicating some aspects of these theorems we set down a definition. A complex-valued variate w is called *complex normal* with mean 0 and variance σ^2 when its real and imaginary parts are independent normal variates with mean 0 and variance $\sigma^2/2$. The density function of w is proportional to $\exp\{-|w|^2/\sigma^2\}$. The variate $|w|^2$ is exponential with mean σ^2 in this case.

In the case that the series $X(t)$ is stationary, with finite second-order moments, and mixing (that is, well-separated values are only weakly dependent) the finite Fourier transform has the following useful asymptotic properties as $T \to \infty$:

a) $d_X^{(T)}(0) - Tc_X$ is asymptotically normal with mean 0 and variance $2\pi Tf_{XX}(0)$;

b) for $\lambda \neq 0, \pm\pi, \pm 2\pi, \cdots, d_X^{(T)}(\lambda)$ is asymptotically complex normal with mean 0 and variance $2\pi Tf_{XX}(\lambda)$;

c) for $s^j(T)$, $j = 1, \cdots, J$ integers with $\lambda^j(T) = 2\pi s^j(T)/T \to \lambda \neq 0, \pm\pi, \pm 2\pi, \cdots$ the variates $d_X^{(T)}(\lambda^1(T)), \cdots, d_X^{(T)}(\lambda^J(T))$ are asymptotically independent complex normals with mean 0 and variance $2\pi Tf_{XX}(\lambda)$,

d) for $\lambda \neq 0, \pm\pi, \pm 2\pi, \cdots$ and $U = T/J$ and integer, the variates

$$d_X^{(U)}(\lambda, j) = \sum_{u=0}^{U-1} X(u + jU) \exp\{-i\lambda u\}, \qquad j = 0, \cdots, J - 1$$

are asymptotically independent complex normals with mean 0 and variance $2\pi Uf_{XX}(\lambda)$.

These results are developed in Brillinger [20]. Related results are given in Section XV and proved in the Appendix. Other references include: Leonov and Shiryaev [23], Picinbono [24], Rosenblatt [25], Brillinger [26], Hannan and Thomson [27]. We have seen that $\exp\{i\lambda t\} \, d_X^{(T)}(\lambda)$ may be interpreted as the result of narrow band-pass filtering the series $X(t)$. It follows that the preceding result b) is consistent with the "engineering folk" theorem to the effect that narrow band-pass noise is approximately Gaussian.

Result a) suggests estimating the mean c_X by

$$c_X^{(T)} = T^{-1} \sum_{t=0}^{T-1} X(t)$$

and approximating the distribution of this estimate by a normal distribution with mean 0 and variance $2\pi f_{XX}(0)/T$. Result b) suggests estimating the power spectrum $f_{XX}(\lambda)$ by the *periodogram*

$$I_{XX}^{(T)}(\lambda) = (2\pi T)^{-1} |d_X^{(T)}(\lambda)|^2 \qquad (13)$$

in the case $\lambda \neq 0, \pm 2\pi, \cdots$. We will say more about this statistic later. It is interesting to note, from c) and d), that

asymptotically independent statistics with mean 0 and variance proportional to the power spectrum at frequency λ may be obtained by either computing the Fourier transform at particular distinct frequencies near λ or by computing them at the frequency λ but based on different time domains. We warn the reader that the results a)–d) are asymptotic. They are to be evaluated in the sense that they might prove reasonable approximations in practice when the domain of observation is large and when values of the series well separated in the domain are only weakly dependent.

On a variety of occasions we will *taper* the data before computing its Fourier transform. This means that we take a data window $\phi^{(T)}(t)$ vanishing for $t < 0, t > T - 1$, and compute the transform

$$d_X^{(T)}(\lambda) = \sum_t \phi^{(T)}(t) \exp\{-i\lambda t\} X(t) \qquad (14)$$

for selected values of λ. One intention of tapering is to reduce the interference of neighboring frequency components. If

$$\Phi^{(T)}(\lambda) = \sum_t \phi^{(T)}(t) \exp\{-i\lambda t\}$$

then the Cramér representation (9) shows that (14) may be written

$$d_X^{(T)}(\lambda) = \int \Phi^{(T)}(\lambda - \alpha) \, dZ_X(\alpha). \qquad (15)$$

From what we have just said, we will want to choose $\phi^{(T)}(t)$ so that $\Phi^{(T)}(\alpha)$ is concentrated near $\alpha = 0, \pm 2\pi, \cdots$. (One convenient choice of $\phi^{(T)}(t)$ takes the form $\phi(t/T)$ where $\phi(u) = 0$ for $u < 0, u \geqslant 1$.) The asymptotic effect of tapering may be seen to be to replace the variance in b) by $2\pi \sum \phi^{(T)}(t)^2 f_{XX}(\lambda)$.

Hannan and Thomson [27] investigate the asymptotic distribution of the Fourier transform of tapered data in a case where $f_{XX}(\lambda)$ depends on T in a particular manner. The hope is to obtain better approximations to the distribution.

V. ESTIMATION OF THE POWER SPECTRUM

In the previous section, we mentioned the periodogram, $I_{XX}^{(T)}(\lambda)$, as a possible estimate of the power spectrum $f_{XX}(\lambda)$ in the case that $\lambda \neq 0, \pm 2\pi, \cdots$. If result b) holds true, then $I_{XX}^{(T)}(\lambda)$, being a continuous function of $d_X^{(T)}(\lambda)$, will be distributed asymptotically as $|w|^2$, where w is a complex normal variate with mean 0 and variance $f_{XX}(\lambda)$. That is $I_{XX}^{(T)}(\lambda)$ will be distributed asymptotically as an exponential variate with mean $f_{XX}(\lambda)$. From the practical standpoint this is interesting, but not satisfactory. It suggests that no matter how large the sample size T is, the variate $I_{XX}^{(T)}(\lambda)$ will tend to be distributed about $f_{XX}(\lambda)$ with an appreciable scatter. Luckily, results c) and d) suggest means around this difficulty. Following c), the variates $I_{XX}^{(T)}(\lambda^j(T)), j = 1, \cdots, J$ are distributed asymptotically as independent exponential variates with mean $f_{XX}(\lambda)$. Their average

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^J I_{XX}^{(T)}(\lambda^j(T)) \qquad (16)$$

will be distributed asymptotically as the average of J independent exponential variates having mean $f_{XX}(\lambda)$. That is, it will be distributed as

$$f_{XX}(\lambda) \chi_{2J}^2 / 2J \qquad (17)$$

where χ_{2J}^2 denotes a chi-squared variate with $2J$ degrees of freedom. The variance of the variate (17) is

$$f_{XX}(\lambda)^2 / J = f_{XX}(\lambda)^2 U/T \qquad (18)$$

if $U = T/J$. By choice of J the experimenter can seek to obtain an estimate of which the sampling fluctuations are small enough for his needs. From the standpoint of practice, it seems to be useful to compute the estimate (16) for a number of values of J. This allows us to tailor the choice of J to the situation at hand and even to use different values of J for different frequency ranges. Result d) suggests our consideration of the estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=0}^{J-1} (2\pi U)^{-1} |d_X^{(U)}(\lambda, j)|^2. \qquad (19)$$

It too will have the asymptotic distribution (17) with variance (18).

We must note that it is not sensible to take J in (16) and (19) arbitrarily large as the preceding arguments might have suggested. It may be seen from (15) that

$$E I_{XX}^{(T)}(\lambda) = \int_{-\pi}^{\pi} F_T(\lambda - \alpha) f_{XX}(\alpha) \, d\alpha + F_T(\lambda) c_X^2 \qquad (20)$$

where

$$F_T(\lambda) = (2\pi T)^{-1} \left| \frac{\sin \dfrac{T\lambda}{2}}{\sin \dfrac{\lambda}{2}} \right|^2$$

is the Fejér kernel. This kernel, or frequency window, is nonnegative, integrates to 1, and has most of its mass in the interval $(-2\pi/T, 2\pi/T)$. The term in c_X^2 may be neglected for $\lambda \neq 0$, $\pm 2\pi, \cdots$ and T large. From (16) and (20) we now see that

$$E f_{XX}^{(T)}(\lambda) \doteq \int_{-\pi}^{\pi} J^{-1} \sum_{j=1}^J F_T(\lambda^j(T) - \alpha) f_{XX}(\alpha) \, d\alpha. \qquad (21)$$

If we are averaging J periodogram values at frequencies $2\pi/T$ apart and centered at λ, then the bandwidth of the kernel of (21) will be approximately $4\pi J/T$. If J is large and $f_{XX}(\alpha)$ varies substantially in the interval $-2\pi J/T < \alpha - \lambda < 2\pi J/T$, then the value of (21) can be very far from the desired $f_{XX}(\lambda)$. In practice we will seek to have J large so that the estimate is reasonably stable, but not so large that it has appreciable bias. This same remark applies to the estimate (19). Parzen [28] constructed a class of estimates such that $E f_{XX}^{(T)}(\lambda) \to f_{XX}(\lambda)$ and var $f_{XX}^{(T)}(\lambda) \to 0$. These estimates have an asymptotic distribution that is normal, rather than χ^2, Rosenblatt [29]. Using the notation preceding these estimates correspond to having J depend on T in such a way that $J_T \to \infty$, but not so large that $J_T/T \to 0$ as $T \to \infty$.

Estimates of the power spectrum have proved useful; i) as simple descriptive statistics, ii) in informal testing and discrimination, iii) in the estimation of unknown parameters, and iv) in the search for hidden periodicities. As an example of i), we mention their use in the description of the color of an object, Wright [30]. In connection with ii) we mention the estimation of the spectrum of the seismic record of an event in attempt to see if the event was an earthquake or a nuclear explo-

sion, Carpenter [31], Lampert *et al.* [32]. In case iii), we mention that Munk and MacDonald [33] derived estimates of the fundamental parameters of the rotation of the Earth from the periodogram. Turning to iv), we remind the reader that the original problem that led to the definition of the power spectrum, was that of the search for hidden periodicities. As a modern example, we mention the examination of spectral estimates for the periods of the fundamental vibrations of the Earth, MacDonald and Ness [34].

VI. OTHER ESTIMATES OF THE POWER SPECTRUM

We begin by mentioning minor modifications that can be made to the estimates of Section V. The periodograms of (16) may be computed at frequencies other than those of the form $2\pi s/T$, s an integer, and they may be weighted unequally. The periodograms of the estimate (19) may be based on overlapping stretches of data. The asymptotic distributions are not so simple when these modifications are made, but the estimate is often improved. The estimate (19) has another interpretation. We saw in Section IV that $\exp\{i\lambda t\} d_X^{(U)}(\lambda, j)$ might be interpreted as the output of a narrow band-pass filter centered at λ. This suggests that (19) is essentially the first power spectral estimate widely employed in practice, the average of the squared output of a narrow band-pass filter (Wegel and Moore [35]). We next turn to a discussion of some spectral estimates of quite different character.

We saw in Section III that if the series $X(t)$ was passed through a linear filter with transfer function $A(\lambda)$, then the output series $Y(t)$ had power spectrum given by $f_{YY}(\lambda) = |A(\lambda)|^2 f_{XX}(\lambda)$. In Section V, we saw that the estimates (16), (19) could have substantial bias were there appreciable variation in the value of the population power spectrum. These remarks suggest a means of constructing an improved estimate, namely: we use our knowledge of the situation at hand to devise a filter, with transfer function $A(\lambda)$, such that the output series $Y(t)$ has spectrum nearer to being constant. We then estimate the power spectrum of the filtered series in the manner of Section V and take $|A(\lambda)|^{-2} f_{YY}^{(T)}(\lambda)$ as our estimate of $f_{XX}(\lambda)$. This procedure is called spectral estimation by prewhitening and is due to Tukey (see Panofsky and McCormick [9]). We mention that in many situations we will be content to just examine $f_{YY}^{(T)}(\lambda)$. This would be necessary were $A(\lambda) = 0$.

One useful means of determining an $A(\lambda)$ is to fit an autoregressive scheme to the data by least squares. That is, for some K, choose $\hat{a}(1), \cdots, \hat{a}(K)$ to minimize

$$\sum [X(t) + a(1) X(t-1) + \cdots + a(K) X(t-K)]^2$$

where the summation extends over the available data. In this case $\hat{A}(\lambda) = 1 + \hat{a}(1) \exp\{-i\lambda\} + \cdots + \hat{a}(K) \exp\{-i\lambda K\}$. An algorithm for efficient computation of the $\hat{a}(u)$ is given in Wiener [36, p. 136]. This procedure should prove especially effective when the series $X(t)$ is near to being an autoregressive scheme of order K. Related procedures are discussed in Grenander and Rosenblatt [37, p. 270], Parzen [38], Lacoss [39], and Burg [40]. Berk [41] discusses the asymptotic distribution of the estimate $|\hat{A}(\lambda)|^{-2}(2\pi T)^{-1} \sum [X(t) + \hat{a}(1) X(t-1) + \cdots + \hat{a}(K) X(t-K)]^2$. Its asymptotic variance is shown to be (18) with $U = 2K$.

Pisarenko [42] has proposed a broad class of estimates including the high resolution estimate of Capon [43] as a particular case. Suppose $\hat{\Sigma}$ is an estimate of the covariance matrix

of the variate

$$\begin{bmatrix} X(1) \\ \vdots \\ X(U) \end{bmatrix}$$

determined from the sample values $X(0), \cdots, X(T-1)$. Suppose $\hat{\mu}_u$, $\hat{\alpha}_u$, $u = 1, \cdots, U$ are the latent roots and vectors of $\hat{\Sigma}$. Suppose $H(\mu)$, $0 < \mu < \infty$, is a strictly monotonic function with inverse $h(\cdot)$. Pisarenko proposed the estimate

$$h\left(\sum_{u=1}^{U} H(\hat{\mu}_u)(2\pi U)^{-1} \left| \sum_{j=1}^{U} \hat{\alpha}_{uj} \exp\{-i\lambda j\} \right|^2 \right). \quad (22)$$

He presents an argument indicating that the asymptotic variance of this estimate is also (18). The hope is that it is less biased. Its character is that of a nonlinear average of periodogram values in contrast to the simple average of (16) and (19). The estimates (16) and (19) essentially correspond to the case $H(\mu) = \mu$. The high resolution estimate of Capon [43] corresponds to $H(\mu) = \mu^{-1}$.

The autoregressive estimate, the high-resolution estimate and the Pisarenko estimates are not likely to be better than an ordinary spectral estimate involving steps of prewhitening, tapering, naive spectral estimation and recoloring. They are probably better than a naive spectral estimate for a series that is a sum of sine waves and noise.

VII. FINITE PARAMETER MODELS

Sometimes a situation arises in which we feel that the form of the power spectrum is known except for the value of a finite dimensional parameter θ. For example existing theory may suggest that the series $X(t)$ is generated by the mixed moving average autoregressive scheme

$$X(t) + a(1)X(t-1) + \cdots + a(K)X(t-K) = \epsilon(t) + b(1)\epsilon(t-1)$$
$$+ \cdots + b(L)\epsilon(t-L) \quad (23)$$

where U, V are nonnegative integers and $\epsilon(t)$ is a series of independent variates with mean 0 and variance σ^2. The power spectrum of this series is

$$f_{XX}(\lambda; \theta) = \frac{\sigma^2}{2\pi} \frac{|1 + b(1) \exp\{-i\lambda\} + \cdots + b(L) \exp\{-i\lambda L\}|^2}{|1 + a(1) \exp\{-i\lambda\} + \cdots + a(K) \exp\{-i\lambda K\}|^2}$$
$$(24)$$

with $\theta = \sigma^2$, $a(1), \cdots, a(K)$, $b(1), \cdots, b(L)$. A number of procedures have been suggested for estimating the parameters of the model (23), see Hannan [44] and Anderson [45], for example.

The following procedure is useful in situations more general than the above. It is a slight modification of a procedure of Whittle [46]. Choose as an estimate of θ the value that maximizes

$$\prod_{0 < s < T/2} f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1} \exp\left\{-I_{XX}^{(T)}\left(\frac{2\pi s}{T}\right) f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1}\right\}.$$
$$(25)$$

Expression (25) is the likelihood corresponding to the assumption that the periodogram values $I_{XX}^{(T)}(2\pi s/T)$, $0 < s < T/2$, are independent exponential variates with means $f_{XX}(2\pi s/T; \theta)$, $0 < s < T/2$, respectively. Under regularity conditions we can show that this estimate, $\hat{\theta}$, is asymptotically normal with mean

θ and covariance matrix $2\pi T^{-1}A^{-1}(A + B)A^{-1}$ where; if $\nabla f_{XX}(\lambda; \theta)$ is the gradient vector with respect to θ and f_{XXXX} the 4th order cumulant spectrum (see Section XV)

$$A = \int_0^\pi \nabla f_{XX}(\alpha; \theta) \cdot \nabla f_{XX}(\alpha; \theta) f_{XX}(\alpha; \theta)^{-2} \, d\alpha$$

$$B = \int_0^\pi \int_0^\pi \nabla f_{XX}(\alpha; \theta) \cdot \nabla f_{XX}(\beta; \theta) f_{XX}(\alpha; \theta)^{-2} f_{XX}(\beta; \theta)^{-2}$$
$$\cdot f_{XXXX}(\alpha, -\alpha, -\beta) \, d\alpha \, d\beta.$$

We may carry out the maximization of (25) by a number of computer algorithms, see the discussion in Chambers [47]. In [48], we used the method of scoring. Other papers investigating estimates of this type are Whittle [49], Walker [50], and Dzaparidze [51].

The power spectrum itself may now be estimated by $f_{XX}(\lambda; \hat\theta)$. This estimate will be asymptotically normal with mean $f_{XX}(\lambda; \theta)$ and variance $2\pi T^{-1} \nabla f_{XX}(\lambda; \theta)^T A^{-1}(A + B) \cdot A^{-1} \nabla f_{XX}(\lambda; \theta)$ following the preceding asymptotic normal distribution for θ. In the case that we model the series by an autoregressive scheme and proceed in the same way, the estimate $f_{XX}(\lambda; \hat\theta)$ has the character of the autoregressive estimate of the previous section.

VIII. Linear Models

In some circumstances we may find ourselves considering a linear time invariant model of the form

$$X(t) = \mu + \sum_{u=-\infty}^{\infty} a(t - u)S(u) + \epsilon(t) \qquad (26)$$

where the values $X(t)$, $S(t)$, $t = 0, 1, \cdots, T - 1$ are given, $\epsilon(t)$ is an unknown stationary error series with mean 0 and power spectrum $f_{\epsilon\epsilon}(\lambda)$, the $a(u)$ are unknown coefficients, μ is an unknown parameter, and $S(t)$ is a fixed function. For example, we might consider the linear trend model

$$X(t) = \mu + at + \epsilon(t)$$

with μ and a unknown, and be interested in estimating $f_{\epsilon\epsilon}(\lambda)$. Or we might have taken $S(t)$ to be the input series to a linear filter with unknown impulse-response function $a(u)$, $u = 0$, $\pm 1, \cdots$ in an attempt to identify the system, that is, to estimate the transfer function $A(\lambda) = \sum a(u) \exp\{-i\lambda u\}$ and the $a(u)$. The model (26) for the series $X(t)$ differs in an important way from the previous models of this paper. The series $X(t)$ is not generally stationary, because $EX(t) = \mu + \sum a(t - u)S(u)$.

Estimates of the preceding parameters may be constructed as follows: define

$$d_X^{(T)}(\lambda) = \sum_{t=0}^{T-1} X(t) \exp\{-i\lambda t\}$$

with similar definitions for $d_S^{(T)}(\lambda)$, $d_\epsilon^{(T)}(\lambda)$. Then (26) leads to the approximate relationship

$$d_X^{(T)}(\lambda) \doteq \mu \sum_{t=0}^{T-1} \exp\{-i\lambda t\} + A(\lambda)d_S^{(T)}(\lambda) + d_\epsilon^{(T)}(\lambda). \quad (27)$$

Suppose $\lambda^1(T), \cdots, \lambda^j(T) \doteq \lambda$ are as in Section IV. Then

$$d_X^{(T)}(\lambda^j(T)) \doteq A(\lambda)d_S^{(T)}(\lambda^j(T)) + d_\epsilon^{(T)}(\lambda^j(T)) \qquad (28)$$

for $j = 1, \cdots, J$. Following b) of Section IV, the $d_\epsilon^{(T)}(\lambda^j(T))$ are, for large T, approximately independent complex normal variates with mean 0 and variance $2\pi T f_{\epsilon\epsilon}(\lambda)$. The approximate model (28) is seen to take the form of linear regression. The results of linear least-squares theory now suggest our consideration of the estimates,

$$A^{(T)}(\lambda) = f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} \qquad (29)$$

and

$$f_{\epsilon\epsilon}^{(T)}(\lambda) = f_{XX}^{(T)}(\lambda) - f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} f_{SX}^{(T)}(\lambda)$$

where

$$f_{SX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} (2\pi T)^{-1} d_S^{(T)}(\lambda^j(T)) \overline{d_X^{(T)}(\lambda^j(T))}$$

with similar definitions for $f_{XS}^{(T)}$, $f_{XX}^{(T)}$, $f_{SS}^{(T)}$. The impulse response could be estimated by an expression such as

$$a^{(T)}(u) = P^{-1} \sum_{p=0}^{P-1} A^{(T)}\left(\frac{2\pi p}{P}\right) \exp\left\{\frac{-i2\pi pu}{P}\right\}$$

for some integer P. In some circumstances it may be appropriate to taper the data prior to computing the Fourier transform. In others it might make sense to base the Fourier transforms on disjoint stretches of data in the manner of d) of Section IV.

Under regularity conditions the estimate $A^{(T)}(\lambda)$ may be shown to be asymptotically complex normal with mean $A(\lambda)$ and variance $J^{-1} f_{\epsilon\epsilon}(\lambda) f_{SS}(\lambda)^{-1}$ (see [20]). The degree of fit of the model (26) at frequency λ may be measured by the sample coherence function

$$|R_{XS}^{(T)}(\lambda)|^2 = |f_{XS}^{(T)}(\lambda)|^2 / [f_{SS}^{(T)}(\lambda) f_{XX}^{(T)}(\lambda)]$$

satisfying

$$f_{\epsilon\epsilon}^{(T)}(\lambda) = [1 - |R_{XS}^{(T)}(\lambda)|^2] f_{XX}^{(T)}(\lambda).$$

This function provides a time series analog of the squared coefficient of correlation of two variates (see Koopmans [52]).

The procedure of prefiltering is often essential in the estimation of the parameters of the model (26). Consider a common relationship in which the series $X(t)$ is essentially a delayed version of the series $S(t)$, namely

$$X(t) = \alpha S(t - v) + \epsilon(t)$$

for some v. In this case

$$A(\lambda) = \alpha \exp\{-i\lambda v\},$$
$$d_X^{(T)}(\lambda^j(T)) = \alpha \exp\{-i\lambda^j(T)v\} d_S^{(T)}(\lambda^j(T)) + d_\epsilon^{(T)}(\lambda^j(T))$$

and

$$f_{XS}^{(T)}(\lambda) = \alpha J^{-1} \sum_j \exp\{-i\lambda^j(T)v\} I_{SS}^{(T)}(\lambda^j(T))$$
$$+ f_{\epsilon S}^{(T)}(\lambda). \quad (30)$$

If v is large, the complex exponential fluctuates rapidly about 0 as j changes and the first term on the right-hand side of (30) may be near 0 instead of the desired $\alpha \exp\{-i\lambda v\} f_{SS}^{(T)}(\lambda)$. A useful prefiltering for this situation is to estimate v by $\hat v$, the lag that maximizes the magnitude of the sample cross-covariance function, and then to carry out the spectral computations

on the data $X(t)$, $S(t - \hat{v})$, see Akaike and Yamanouchi [53] and Tick [54]. In general, one should prefilter the $X(t)$ series or the $S(t)$ series or both, so that the relationship between the filtered series is as near to being instantaneous as is possible.

The most important use of the calculations we have described is in the identification of linear systems. It used to be the case that the transfer function of a linear system was estimated by probing the system with pure sine waves in a succession of experiments. Expression (29) shows, however, that we can estimate the transfer function, for all λ, by simply employing a single input series $S(t)$ such that $f_{SS}^{(T)}(\lambda) \neq 0$.

In some situations we may have reason to believe that the system (26) is *realizable* that is $a(u) = 0$ for $u < 0$. The factorization techniques of Wiener [36] may be paralleled on the data in order to obtain estimates of $A(\lambda)$, $a(u)$ appropriate to this case, see Bhansali [55]. In Section IX, we will discuss a model like (26), but for the case of stochastic $S(t)$.

Another useful linear model is

$$X(t) = \theta_1 \phi_1(t) + \cdots + \theta_K \phi_K(t) + \epsilon(t)$$

with $\phi_1(t), \cdots, \phi_K(t)$ given functions and $\theta_1, \cdots, \theta_K$ unknown. The estimation of these unknowns and $f_{\epsilon\epsilon}(\lambda)$ is considered in Hannan [44] and Anderson [45]. This model allows us to handle trends and seasonal effects.

Yet another useful model is

$$X(t) = \mu + \rho_1 \sin(\theta_1 t + \alpha_1) + \cdots + \rho_K \sin(\theta_K t + \alpha_K) + \epsilon(t)$$

with $\mu, \rho_1, \theta_1, \alpha_1, \cdots, \rho_K, \theta_K, \alpha_K$ unknown. The estimation of these unknowns and $f_{\epsilon\epsilon}(\lambda)$ is considered in Whittle [49]. It allows us to handle hidden periodicities.

IX. VECTOR-VALUED CONTINUOUS SPATIAL SERIES

In this section we move on from a consideration of real-valued discrete time series to series with a more complicated domain, namely p-dimensional Euclidean space, and with a more complicated range, namely r-dimensional Euclidean space. This step will allow us to consider data such as: that received by an array of antennas or seismometers, picture or TV, holographic, turbulent field.

Provided we set down our notation judiciously, the changes involved are not dramatic. The notation that we shall adopt includes the following: boldface letters such as X, a, A will denote vectors and matrices. A^T will denote the transpose of a matrix A, tr A will denote its trace, det A will denote its determinant. EX will denote the vector whose entries are the expected values of the corresponding entries of the vector-valued variate X. cov $\{X, Y\} = E\{(X - EX)\overline{(Y - EY)}^T\}$ will denote the covariance matrix of the two vector-valued variates X, Y (that may have complex entries). t, u, λ will lie in p-dimensional Euclidean space, R^p, with

$$t = (t_1, \cdots, t_p) \qquad dt = dt_1 \cdots dt_p$$
$$u = (u_1, \cdots, u_p) \qquad du = du_1 \cdots du_p$$
$$\lambda = (\lambda_1, \cdots, \lambda_p) \qquad d\lambda = d\lambda_1 \cdots d\lambda_p$$
$$\langle \lambda, t \rangle = \lambda_1 t_1 + \cdots + \lambda_p t_p$$
$$\langle \lambda, u \rangle = \lambda_1 u_1 + \cdots + \lambda_p u_p$$
$$|u| = (u_1^2 + \cdots + u_p^2)^{1/2}$$
$$|\lambda| = (\lambda_1^2 + \cdots + \lambda_p^2)^{1/2}.$$

The limits of integrals will be from $-\infty$ to ∞, unless indicated otherwise.

We will proceed by paralleling the development of Sections III and IV. Suppose that we are interested in analyzing measurements made simultaneously on r series of interest at location t, for all locations in some subset of the hypercube $0 < t_1, \cdots, t_p < T$. Suppose that we are prepared to model the measurements by the corresponding values of a realization of an r vector-valued stationary continuous spatial series $X(t)$, $t \in R^p$. We define the *mean*

$$c_X = EX(t)$$

the *autocovariance function*

$$c_{XX}(u) = \text{cov}\{X(t + u), X(t)\}$$

and the *spectral density matrix*

$$f_{XX}(\lambda) = (2\pi)^{-p} \int \exp\{-i\langle \lambda, u \rangle\} c_{XX}(u)\, du, \qquad \lambda \in R^p \quad (31)$$

in the case that the integral exists. (The integral will exist when well-separated values of the series are sufficiently weakly dependent.) The inverse of the relationship (31) is

$$c_{XX}(u) = \int \exp\{i\langle \lambda, \alpha \rangle\} f_{XX}(\alpha)\, d\alpha. \qquad (32)$$

Let

$$X(t) \rightarrow Y(t) = \int a(t - u) X(u)\, du$$

be a linear filter carrying the r vector-valued series $X(t)$ into the s vector-valued series $Y(t)$. Let

$$A(\lambda) = \int a(u) \exp\{-i\langle \lambda, u \rangle\}\, du$$

denote the transfer function of this filter. Then the spectral density matrix of the series $Y(t)$ may be seen to be

$$f_{YY}(\lambda) = A(\lambda) f_{XX}(\lambda) \overline{A(\lambda)}^T. \qquad (33)$$

As in Section III, expressions (32) and (33) may be combined to see that the entry in row j, column k of the matrix $f_{XX}(\lambda)$ may be interpreted as the covariance of the series resulting from passing the jth and kth components of $X(t)$ through narrow band-pass filters with transfer functions $A(\alpha) = \delta(\alpha - \lambda)$.

The series has a Cramér representation

$$X(t) = \int \exp\{i\langle \alpha, t \rangle\}\, dZ_X(\alpha)$$

where $Z_X(\lambda)$ is an r vector-valued random function with the properties

$$E\, dZ_X(\lambda) = \delta(\lambda) c_X\, d\lambda$$

$$\text{cov}\{dZ_X(\lambda), dZ_X(\mu)\} = \delta(\lambda - \mu) f_{XX}(\lambda)\, d\lambda\, d\mu.$$

If $Y(t)$ is the filtered version of $X(t)$, then it has Cramér representation

$$Y(t) = \int \exp\{i\langle \alpha, t \rangle\} A(\alpha)\, dZ_X(\alpha).$$

We turn to a discussion of useful computations when values of the series $X(t)$ are available for t in some subset of the hypercube $0 < t_1, \cdots, t_p < T$. Let $\phi^{(T)}(t)$ be a data window whose support (that is the region of locations where $\phi^{(T)}(t) \neq 0$) is the region of observation of $X(t)$. (We might take $\phi^{(T)}(t)$ of the form $\phi(t/T)$ where $\phi(t) = 0$ outside $0 < t_1, \cdots, t_p < 1$.) We consider the Fourier transform

$$d_X^{(T)}(\lambda) = \int X(t)\phi^{(T)}(t) \exp\{-i\langle\lambda, t\rangle\} \, dt$$

based on the observed sample values.

Before indicating an approximate large sample distribution for $d_X^{(T)}(\lambda)$, we must first define the complex multivariate normal distribution and the complex Wishart distribution. We say that a vector-valued variate X, with complex entries, is *multivariate complex normal* with mean $\mathbf{0}$ and covariance matrix Σ when it has probability density proportional to $\exp\{-\bar{X}^\tau \Sigma^{-1} X\}$. We shall say that a matrix-valued variate is *complex Wishart* with n degrees of freedom and parameter Σ when it has the form $X_1\bar{X}_1^\tau + \cdots + X_n\bar{X}_n^\tau$, where $X_1, \cdots, X_n$ are independent multivariate complex normal variates with mean $\mathbf{0}$ and covariance matrix Σ. In the one dimensional case, the complex Wishart with n degrees of freedom is a multiple of a chi-squared variate with $2n$ degrees of freedom.

In the case that well-separated values of the series $X(t)$ are only weakly dependent, the $d_X^{(T)}(\lambda)$ have useful asymptotic properties as $T \to \infty$. These include:

a') $d_X^{(T)}(0)$ is asymptotically multivariate normal with mean $\int \phi^{(T)}(t) \, dt \, c_X$ and covariance matrix $(2\pi)^p \int \phi^{(T)}(t)^2 \, dt \, f_{XX}(0)$;

b') for $\lambda \neq 0$, $d_X^{(T)}(\lambda)$ is asymptotically multivariate complex normal with mean $\mathbf{0}$ and covariance matrix

$$(2\pi)^p \int \phi^{(T)}(t)^2 \, dt \, f_{XX}(\lambda);$$

c') for $\lambda^j(T) \to \lambda \neq 0$, with $\lambda^j(T) - \lambda^k(T)$ not tending to 0 too rapidly, $1 \leqslant j < k \leqslant J$, the variates $d_X^{(T)}(\lambda^1(T)), \cdots, d_X^{(T)}(\lambda^J(T))$ are asymptotically independent multivariate complex normal with mean $\mathbf{0}$ and covariance matrix

$$(2\pi)^p \int \phi^{(T)}(t)^2 \, dt \, f_{XX}(\lambda);$$

d') if $\phi_j^{(T)}(t)\phi_k^{(T)}(t) = 0$, for all t, $1 \leqslant j < k \leqslant J$, and if $\lambda \neq 0$ the variates

$$d_X^{(T)}(\lambda, j) = \int X(t)\phi_j^{(T)}(t) \exp\{-i\langle\lambda, t\rangle\} \, dt \qquad (34)$$

$j = 1, \cdots, J$ are asymptotically independent multivariate complex normal with mean $\mathbf{0}$ and respective covariance matrices $(2\pi)^p \int \phi^{(T)}(t, j)^2 \, dt \, f_{XX}(\lambda)$, $j = 1, \cdots, J$.

Specific conditions under which these results hold are given in Section XV. A proof is given in the Appendix.

Results a'), b') are forms of the central limit theorem. In result d') the Fourier transforms are based on values of $X(t)$ over disjoint domains. It is interesting to note, from c') and d') that asymptotically independent statistics may be obtained by either taking the Fourier transform at distinct frequencies or at the same frequency, but over disjoint domains.

Result a') suggests estimating the mean c_X by

$$c_X^{(T)} = \frac{\int X(t)\phi^{(T)}(t) \, dt}{\int \phi^{(T)}(t) \, dt}. \qquad (35)$$

Result b') suggests the consideration of the *periodogram matrix*

$$I_{XX}^{(T)}(\lambda) = (2\pi)^{-p}\left(\int \phi^{(T)}(t)^2 \, dt\right)^{-1} d_X^{(T)}(\lambda)\overline{d_X^{(T)}(\lambda)}^\tau \qquad (36)$$

as an estimate of $f_{XX}(\lambda)$ when $\lambda \neq 0$. From b') its asymptotic distribution is complex Wishart with 1 degree of freedom and parameter $f_{XX}(\lambda)$. This estimate is often inappropriate because of its instability and singularity. Result c') suggests the consideration of the estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda^j(T)) \qquad (37)$$

where J is chosen large enough to obtain acceptable stability, but not so large that the estimate becomes overly biased. From c') the asymptotic distribution of the estimate (37) is complex Wishart with J degrees of freedom and parameter $f_{XX}(\lambda)$. In the case $J = 1$ this asymptotic distribution is that of $f_{XX}(\lambda)\chi_{2J}^2/2J$. Result d') suggests the consideration of the periodogram matrices

$$I_{XX}^{(T)}(\lambda, j) = (2\pi)^{-p}\left(\int \phi_j^{(T)}(t)^2 \, dt\right)^{-1} d_X^{(T)}(\lambda, j)\overline{d_X^{(T)}(\lambda, j)}^\tau \qquad (38)$$

$j = 1, \cdots, J$ as estimates of $f_{XX}(\lambda)$, $\lambda \neq 0$. The estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda, j) \qquad (39)$$

will have as asymptotic distribution J^{-1} times a complex Wishart with J degrees of freedom and parameter $f_{XX}(\lambda)$ following result d'). We could clearly modify the estimates (37), (39) by using a finer spacing of frequencies and by averaging periodograms based on data over nondisjoint domains. The exact asymptotic distributions will not be so simple in these cases.

The method of fitting finite parameter models, described in Section VII, extends directly to this vector-valued situation. Result b') suggests the replacement of the likelihood function (25) by

$$\prod_{0 < s_j < S_j} \det f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1}$$

$$\cdot \exp\left\{-\text{tr } I_{XX}^{(T)}\left(\frac{2\pi s}{T}\right)f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1}\right\} \qquad (40)$$

in this new case for some large values $S_1, \cdots, S_p$ such that there is little power left beyond the cutoff frequency $(2\pi S_1/T, \cdots, 2\pi S_p/T)$. Suppose that $\hat{\theta}$ is the value of θ leading to the maximum of (40). Under regularity conditions, we can show that $\hat{\theta}$ is asymptotically normal with mean θ and covariance matrix $2\pi T^{-1}A^{-1}(A + B)A^{-1}$ where if A_{jk}, B_{jk} are

row j, column k of A, B

$$A_{jk} = \int_0^{2\pi S/T} \operatorname{tr} \frac{\partial f(\alpha)}{\partial \theta_j} f(\alpha)^{-1} \frac{\partial f(\alpha)}{\partial \theta_k} f(\alpha)^{-1} \, d\alpha$$

$$B_{jk} = \int_0^{2\pi S/T} \sum_a \sum_b \sum_c \sum_d C_{abj}(\alpha) C_{cdk}(\beta) f_{abcd}(\alpha, -\alpha, -\beta)$$
$$\cdot \, d\alpha \, d\beta$$

with $C_{abj}(\alpha)$ the entry in row a column b of

$$f(\alpha)^{-1} \frac{\partial f(\alpha)}{\partial \theta_j} f(\alpha)^{-1}.$$

In a number of situations we find ourselves led to consider an $(r + s)$ vector-valued series,

$$\begin{bmatrix} S(t) \\ X(t) \end{bmatrix} \qquad (41)$$

satisfying a linear model of the form

$$E\{X(t) | S(u), u \in R^p\} = \mu + \int a(t - u) S(u) \, du \qquad (42)$$

for some s vector μ and some $s \times r$ matrix-valued function $a(u)$. The model says that the average level of the series $X(t)$ at position t, given the series $S(t)$, is a linear filtered version of the series $S(t)$. If (41) is a stationary series and if $A(\lambda)$ is the transfer function of the filter $a(u)$, then (42) implies

$$c_X = \mu + A(0) c_S \qquad (43)$$

$$f_{XS}(\lambda) = A(\lambda) f_{SS}(\lambda). \qquad (44)$$

If we define the error series $\epsilon(t)$ by

$$\epsilon(t) = X(t) - \mu - \int a(t - u) S(u) \, du$$

then the degree of fit of the model (42) may be measured by the error spectral density

$$f_{\epsilon\epsilon}(\lambda) = f_{XX}(\lambda) - f_{XS}(\lambda) f_{SS}(\lambda)^{-1} f_{SX}(\lambda). \qquad (45)$$

The relationships (43)–(45) suggest the estimates

$$A^{(T)}(\lambda) = f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} \qquad (46)$$

$$\mu^{(T)} = c_X^{(T)} - A^{(T)}(0) c_S^{(T)} \qquad (47)$$

$$f_{\epsilon\epsilon}^{(T)}(\lambda) = f_{XX}^{(T)}(\lambda) - f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} f_{SX}^{(T)}(\lambda) \qquad (48)$$

respectively. The asymptotic distributions of these statistics are given in [26].

If there is a possibility that the matrix $f_{SS}^{(T)}(\lambda)$ might become nearly singular, then we would be better off replacing the estimate (46) by a frequency domain analog of the ridge regression estimate (Hoerl and Kennard [56], Hunt [57]), such as

$$f_{XS}^{(T)}(\lambda) [f_{SS}^{(T)}(\lambda) + kI]^{-1} \qquad (49)$$

for some $k > 0$ and I the identity matrix. This estimate introduces further bias, over what was already present, but it is hoped that its increased stability more than accounts for this. In some circumstances we might choose k to depend on λ and to be matrix-valued.

X. Additional Results in the Spatial Series Case

The results of the previous section have not taken any essential notice of the fact that the argument t of the random function under consideration is multidimensional. We now indicate some new results pertinent to the multidimensional character.

In some situations, we may be prepared to assume that the series $X(t)$, $t \in R^p$, is *isotropic*, that is the autocovariance function $c_{XX}(u) = \operatorname{cov}\{X(t + u), X(t)\}$ is a function of $|u|$ only. In this case the spectral density matrix $f_{XX}(\lambda)$ is also rotationally symmetric, depending only on $|\lambda|$. In fact (see in Bochner and Chandrasekharan [58, p. 69])

$$f_{XX}(\lambda) = (2\pi)^{-p/2} |\lambda|^{(2-p)/2} \int_0^\infty |u|^{p/2}$$
$$\cdot J_{(p-2)/2}(|\lambda||u|) c_{XX}(u) \, d|u| \qquad (50)$$

where $J_k(t)$ is the Bessel function of the first kind of order k. The relationship (50) may be inverted as follows,

$$c_{XX}(u) = (2\pi)^{p/2} |u|^{(2-p)/2} \int_0^\infty |\lambda|^{p/2}$$
$$\cdot J_{(p-2)/2}(|\lambda||u|) f_{XX}(\lambda) \, d|\lambda|.$$

The simplified character of $f_{XX}(\lambda)$ in the isotropic case makes its estimation and display much simpler. We can estimate it by an expression such as

$$J^{-1} \sum_{j=1}^J I_{XX}^{(T)}(\lambda^j(T)) \qquad (51)$$

where the $\lambda^j(T)$ are distinct, but with $|\lambda^j(T)|$ near $|\lambda|$. There are many more $\lambda^j(T)$ with $|\lambda^j(T)|$ near $|\lambda|$ than there are $\lambda^j(T)$ with $\lambda^j(T)$ near λ. It follows that we generally obtain a much better estimate of the spectrum in this case over the estimate in the general case. Also the number of $\lambda^j(T)$ with $|\lambda^j(T)|$ near $|\lambda|$ increases as $|\lambda|$ increases. If follows that the estimate formed will generally be more stable for the frequencies with $|\lambda|$ large. Examples of power spectra estimated in this manner may be found in Mannos [59].

Another different thing that can occur in the general p dimensional case is the definition of marginal processes and marginal spectra. We are presently considering processes $X(t_1, \cdots, t_p)$. Suppose that for some n, $1 \leqslant n < p$, we are interested in the process with $t_{n+1}, \cdots, t_p$ fixed, say at $0, \cdots, 0$. By inspection we see that the marginal process $X(t_1, \cdots, t_n, 0, \cdots, 0)$ has autocovariance function $c_{XX}(u_1, \cdots, u_n, 0, \cdots, 0)$. The spectral density matrix of the marginal process is, therefore,

$$(2\pi)^{-n} \int \cdots \int c_{XX}(u_1, \cdots, u_n, 0, \cdots, 0)$$
$$\cdot \exp\{-i(\lambda_1 u_1 + \cdots + \lambda_n u_n)\} \, du_1 \cdots du_n$$
$$= \int \cdots \int f_{XX}(\lambda_1, \cdots, \lambda_n, \lambda_{n+1}, \cdots, \lambda_p)$$
$$\cdot d\lambda_{n+1} \cdots d\lambda_p.$$

We see that we obtain the spectral density of the marginal process by integrating the complete spectral density. The same

remark applies to the Cramér representation for

$$X(t_1, \cdots, t_n, 0, \cdots, 0) = \int \cdots \int \exp\{i(t_1\lambda_1 + \cdots + t_n\lambda_n)\}$$

$$\cdot \int \cdots \int dZ_X(\lambda_1, \cdots, \lambda_p).$$

Vector-valued series with multidimensional domain are discussed in Hannan [44] and Brillinger [26].

XI. ADDITIONAL RESULTS IN THE VECTOR CASE

In the case that the series $X(t)$ is r vector-valued with $r > 1$, we can describe analogs of the classical procedures of multivariate analysis including for example; i) partial correlation, ii) principal component analysis, iii) canonical correlation analysis, iv) cluster analysis, v) discriminant analysis, vi) multivariate analysis of variance, and vii) simultaneous equations. These analogs proceed from c') or d') of earlier section. The procedures listed are often developed for samples from multivariate normal distributions. We obtain the time series procedure by identifying the $d_X^{(T)}(\lambda^j(T))$, $j = 1, \cdots, J$ or $d_X^{(T)}(\lambda, j)$, $j = 0, \cdots, J - 1$ with independent multivariate normals having mean $\mathbf{0}$ and covariance matrix $(2\pi)^p \int \phi^{(T)}(t)^2 \, dt \, f_{XX}(\lambda)$ and substituting into the formulas developed for the classical situation. For example, stationary time series analogs of correlation coefficients are provided by the

$$R_{jk}(\lambda) = f_{jk}(\lambda)/\sqrt{f_{jj}(\lambda)f_{kk}(\lambda)}$$

$$\sim \text{cov}\{d_j^{(T)}(\lambda), d_k^{(T)}(\lambda)\}/\sqrt{\text{var}\, d_j^{(T)}(\lambda)\, \text{var}\, d_k^{(T)}(\lambda)}$$

the *coherency* at frequency λ of the jth component with the kth component of $X(t)$, where $f_{jk}(\lambda)$ is the entry in row j, column k of $f_{XX}(\lambda)$ and $d_j^{(T)}(\lambda)$ is the entry in row j of $d_X^{(T)}(\lambda)$ for $j, k = 1, \cdots, r$. The parameter $R_{jk}(\lambda)$ satisfies $0 \leq |R_{jk}(\lambda)| \leq 1$ and is seen to provide a measure of the degree of linear relationship of the series $X_j(t)$ with the series $X_k(t)$ at frequency λ. Its modulus squared, $|R_{jk}(\lambda)|^2$, is called the *coherence*. It may be estimated by

$$R_{jk}^{(T)}(\lambda) = f_{jk}^{(T)}(\lambda)/\sqrt{f_{jj}^{(T)}(\lambda)f_{kk}^{(T)}(\lambda)}$$

where $f_{jk}^{(T)}(\lambda)$ is an estimate of $f_{jk}(\lambda)$.

As time series papers on corresponding multivariate topics, we mention in case i) Tick [60], Granger [61], Goodman [62], Bendat and Piersol [63], Groves and Hannan [64], and Gersch [65]; in case ii) Goodman [66], Brillinger [67], [20], and Priestley *et al.* [68]; in case iii) Brillinger [67], [20], Miyata [69], and Priestley *et al.* [68]; in case iv) Ligett [70]; in case v) Brillinger [20]; in case vi) Brillinger [71]; in case vii) Brillinger and Hatanaka [72], and Hannan and Terrell [73].

Instead of reviewing each of the time series analogs we content ourselves by indicating a form of discriminant analysis that can be carried out in the time series situation. Suppose that a segment of the r vector-valued series $X(t)$ is available and that its spectral density matrix may be any one of $f_i(\lambda)$, $i = 1, \cdots, I$. Suppose that we wish to construct a rule for assigning $X(t)$ to one of the $f_i(\lambda)$.

In the case of a variate U coming from one of I multivariate normal populations with mean $\mathbf{0}$ and covariance matrix Σ_i, $i = 1, \cdots, I$, a common discrimination procedure is to define a

discriminant score

$$-\tfrac{1}{2} \log \det \Sigma_i - \tfrac{1}{2} U^T \Sigma_i^{-1} U$$

for the ith population and then to assign the observation U to the population for which the discriminant score has the highest value (see Rao [74, p. 488]). The discriminant score is essentially the logarithm of the probability density of the ith population.

Result 2) suggests a time series analog for this procedure. If the spectral density of the series $X(t)$ is $f_i(\lambda)$, the log density of $d_X^{(T)}(\lambda)$ is essentially

$$-\log \det f_i(\lambda) - \text{tr}\, I_{XX}^{(T)}(\lambda)f_i(\lambda)^{-1}. \tag{52}$$

This provides a discriminant score for each frequency λ. A more stable score would be provided by the smoothed version

$$-J^{-1} \log \det f_i(\lambda) - \text{tr}\, f_{XX}^{(T)}(\lambda)f_i(\lambda)^{-1}$$

with $f_{XX}^{(T)}(\lambda)$ given by (37) or (39). These scores could be plotted against λ for $i = 1, \cdots, I$ in order to carry out the required discrimination. In the case that the $f_i(\lambda)$ are unknown, their values could be replaced by estimates in (52).

XII. ADDITIONAL RESULTS IN THE CONTINUOUS CASE

In Section IX, we changed to a continuous domain in contrast to the discrete domain we began with in Section III. In many problems, we must deal with both sorts of domains, because while the phenomenon of interest may correspond to a continuous domain, observational and computational considerations may force us to deal with the values of the process for a discrete domain. This occurrence gives rise to the complication of *aliasing*. Let Z denote the set of integers, $Z = 0, \pm1, \cdots$. Suppose $X(t)$, $t \in R^p$, is a stationary continuous spatial series with spectral density matrix $f_{XX}(\lambda)$ and Cramér representation

$$X(t) = \int \exp\{i\langle \alpha, t\rangle\}\, dZ_X(\alpha).$$

Suppose $X(t)$ is observable only for $t \in Z^p$. For these values of t

$$X(t) = \int_{(-\pi, \pi)^p} \exp\{i\langle \alpha, t\rangle\} \sum_{j \in Z^p} dZ_X(\alpha + 2\pi j).$$

This is the Cramér representation of a discrete series with spectral density matrix

$$\sum_{j \in Z^p} f_{XX}(\lambda + 2\pi j).$$

We see that if the series $X(t)$ is observable only for $t \in Z^p$, then there is no way of untangling the frequencies

$$\lambda + 2\pi j, \quad j \in Z^p.$$

These frequencies are called the aliases of the fundamental frequency λ.

XIII. STATIONARY POINT PROCESSES

A variety of problems, such as those of traffic systems, queues, nerve pulses, shot noise, impulse noise, and microscopic theory of gases lead us to data that has the character of times or positions in space at which certain events have occurred. We turn now to indicating how the formulas we have

presented so far in this paper must be modified to apply to data of this new character.

Suppose that we are recording the positions in p-dimensional Euclidean space at which events of r distinct types occur. For $j = 1, \cdots, r$ let $X_j(t) = X_j(t_1, \cdots, t_p)$ denote the number of events of the jth type that occur in the hypercube $(0, t_1] \times \cdots \times (0, t_p]$. Let $dX_j(t)$ denote the number that occur in the small hypercube $(t_1, t_1 + dt_1] \times \cdots \times (t_p, t_p + dt_p]$. Suppose that joint distributions of variates such as $dX(t^1), \cdots, dX(t^k)$ are unaffected by simple translation of $t^1, \cdots t^k$, we then say that $X(t)$ is a stationary point process.

Stationary point process analogs of definitions set down previously include

$$c_X \, dt = E \, dX(t) \tag{53}$$

c_X is called the *mean intensity* of the process,

$$dC_{XX}(u) \, dt = \operatorname{cov} \{dX(t + u), dX(t)\} \tag{54}$$

$$f_{XX}(\lambda) = (2\pi)^{-p} \int \exp \{-i\langle \lambda, u \rangle\} \, dC_{XX}(u) \tag{55}$$

$$X(t) = \int \cdots \int \left[\frac{\exp \{i\lambda_1 t_1\} - 1}{i\lambda_1}\right]$$
$$\cdots \left[\frac{\exp \{i\lambda_p t_p\} - 1}{i\lambda_p}\right]$$
$$\cdot dZ_X(\lambda_1, \cdots, \lambda_p) \tag{56}$$

$$dX(t) = \int \exp \{i\langle \lambda, t \rangle\} \, dZ_X(\lambda) \, dt \tag{57}$$

$$E\{dX(t)|S(u), u \in R^p\} = \left[\mu + \int a(t - u) \, dS(u)\right] dt. \tag{58}$$

This last refers to an $(r + s)$ vector-valued point process. It says that the instantaneous intensity of the series $X(t)$ at position t, given the location of all the points of the process $S(u)$, is a linear translation invariant function of the process $S(u)$. The locations of the points of $X(t)$ are affected by where the points of $S(u)$ are located. We may define here a stationary random measure $d\epsilon(t)$ by

$$d\epsilon(t) = dX(t) - \left[\mu + \int a(t - u) \, dS(u)\right] dt. \tag{59}$$

We next indicate some statistics that it is useful to calculate when the process $X(t)$ has been observed over some region. The Fourier transform is now

$$d_X^{(T)}(\lambda) = \int \phi^{(T)}(t) \exp \{-i\langle \lambda, t \rangle\} \, dX(t) \tag{60}$$

for the data window $\phi^{(T)}(t)$ whose support corresponds to the domain of observation. If $r = 1$ and points occur at the positions $\tau_1, \tau_2, \cdots$, then this last has the form

$$\phi^{(T)}(\tau_1) \exp \{-i\langle \lambda, \tau_1 \rangle\} + \phi^{(T)}(\tau_2) \exp \{-i\langle \lambda, \tau_2 \rangle\} + \cdots.$$

We may compute Fourier transforms for different domains in which case we define

$$d_X^{(T)}(\lambda, j) = \int \phi_j^{(T)}(t) \exp \{-i\langle \lambda, t \rangle\} \, dX(t). \tag{61}$$

The change in going from the case of spatial series to the case of point processes is seen to be the replacement of $X(t) \, dt$ by $dX(t)$. In the case that well-separated increments of the process are only weakly dependent, the results a')–d') of Section IX hold without further redefinition.

References to the theory of stationary point processes include: Cox and Lewis [75], Brillinger [76], Daley and Vere-Jones [77], and Fisher [78]. We remark that the material of this section applies equally to the case in which $dX(t)$ is a general stationary random measure, for example with $p, r = 1$, we might take $dX(t)$ to be the amount of energy released by earthquakes in the time interval $(t, t + dt)$. In the next section we indicate some results that do take note of the specific character of a point process.

XIV. New Things in the Point Process Case

In the case of a point process, the parameters c_X, $C_{XX}(u)$ have interpretations further to their definitions (53), (54). Suppose that the process is *orderly*, that is the probability that a small region contains more than one point is very small. Then, for small dt

$$c_j \, dt = E \, dX_j(t) \doteq \Pr [\text{there is an event of type } j \text{ in } (t, t + dt]].$$

It follows that c_j may be interpreted as the intensity with which points of type j are occurring. Likewise, for $u \neq 0$

$$dC_{jk}(u) \, dt = \operatorname{cov} \{dX_j(t + u), dX_k(t)\}$$
$$\doteq \Pr [\text{there is an event of type } j \text{ in}$$
$$(t + u, t + u + du] \text{ and an event of type } k \text{ in}$$
$$(t, t + dt]] - c_j c_k \, dt \, du.$$

It follows that

$$\frac{dC_{jk}(u) + c_j c_k \, du}{c_k} = \Pr [\text{event of sort } j \text{ in}$$
$$(t + u, t + u + du] \text{ given an event}$$
$$\text{of sort } k \text{ in } (t, t + dt]]. \tag{62}$$

In the case that the processes $X_j(t)$ and $X_k(t)$ are independent, expression (62) is equal to $c_j du$.

If the derivative $c_{jk}(u) = dC_{jk}(u)/du$ exists for $u \neq 0$ it is called the *cross-covariance density* of the two processes in the case $j \neq k$ and the *autocovariance density* in the case $j = k$. For many processes

$$dC_{jj}(u) = c_j \delta(u) \, du + c_{jj}(u) \, du$$

and so the power spectrum of the process $X_j(t)$ is given by

$$f_{jj}(\lambda) = (2\pi)^{-p} \left[c_j + \int \exp \{-i\langle \lambda, u \rangle\} c_{jj}(u) \, du\right].$$

For a Poisson process $c_{jj}(u) = 0$ and so $f_{XX}(\lambda) = (2\pi)^{-p} c_X$.

The parameter $(2\pi)^p f_{XX}(0)/c_X$ is useful in the classification of real-valued point processes. From 1)

$$\operatorname{var} X(T, \cdots, T) \sim (2\pi)^p T^p f_{XX}(0).$$

It follows that, for large T, $(2\pi)^p f_{XX}(0)/c_X$ is the ratio of the variance of the number of points in the hypercube $(0, T]^p$ for the process $X(t)$ to the variance of the number of points in the same hypercube for a Poisson process with the same intensity c_X. For this reason we say that the process $X(t)$ is *underdispersed* or *clustered* if the ratio is greater than 1 and *overdispersed* if the ratio is less than 1.

The estimation procedure described in Section XI for models with a finite number of parameters is especially useful in the

point process case as, typically, convenient time domain estimation procedures do not exist at all. Results of applying such a procedure are indicated in [79].

XV. STATIONARY RANDOM SCHWARTZ DISTRIBUTIONS

In this section, we present the theory of Schwartz distributions (or generalized functions) needed to develop properties of the Fourier transforms of random Schwartz distributions. These last are important as they contain the processes discussed so far in this paper as particular cases. In addition they contain other interesting processes as particular cases, such as processes whose components are a combination of the processes discussed so far and such as the processes with stationary increments that are useful in the study of turbulence, see Yaglom [80]. A further advantage of this abstract approach is that the assumptions needed to develop results are cut back to essentials. References to the theory of Schwartz distributions include Schwartz [81] and Papoulis [82].

Let $\mathfrak{D}$ denote the space of infinitely differentiable functions on R^p with compact support. Let $\mathfrak{S}$ denote the space of infinitely differentiable functions on R^p with rapid decrease, that is such that if $\phi^{(q)}(t)$ denotes a derivative of order q then

$$\lim_{|t| \to \infty} (1 + |t|)^n \phi^{(q)}(t) \to 0 \quad \text{for all } n, q.$$

A continuous linear functional on $\mathfrak{D}$ is called a *Schwartz distribution* or *generalized function*. The Dirac delta function that we have been using throughout the paper is an example. A continuous linear functional on $\mathfrak{S}$ is called a *tempered distribution*.

Suppose now that a random experiment is being carried out, the possible results of which are continuous linear maps X from $\mathfrak{D}$ to $L^2(P)$, the space of square integrable functions for a probability measure P. Suppose that r of these maps are collected into an r vector, $X(\phi)$. We call $X(\phi)$ an r vector-valued *random Schwartz distribution*. It is possible to talk about things such as $E\, X(\phi)$, cov $\{X(\phi), X(\psi)\}$ in this case. An important family of transformations on $\mathfrak{D}$ consists of the shifts S^u defined by $S^u \phi(t) = \phi(t + u)$, $t, u \in R^p$. The random Schwartz distribution is called *wide-sense stationary* when

$$E\, X(S^u \phi) = E\, X(\phi)$$

$$\text{cov } \{X(S^u \phi), X(S^u \psi)\} = \text{cov } \{X(\phi), X(\psi)\}$$

for all $u \in R^p$ and $\phi, \psi \in \mathfrak{D}$. It is called *strictly stationary* when all the distributions of finite numbers of values are invariant under the shifts.

Let us denote the convolution of two functions $\phi, \psi \in \mathfrak{D}$ by

$$\phi * \psi(t) = \int \phi(t - u) \, \overline{\psi(u)} \, du$$

and the Fourier transform of a function in $\mathfrak{S}$ by the corresponding capital letter

$$\Phi(\lambda) = \int \phi(u) \exp \{-i \langle \lambda, u \rangle\} \, du$$

then we can set down the following Theorem.

Theorem 1: (Ito [83], Yaglom [80].) If $X(\phi)$, $\phi \in \mathfrak{D}$ is a wide-sense stationary random Schwartz distribution, then

$$E\, X(\phi) = c_X \int \phi(t) \, dt \qquad (63)$$

$$\text{cov } \{X(\phi), X(\psi)\} = c_{XX}(\phi * \psi^-) \qquad (64)$$

$$= \int \Phi(-\alpha) \overline{\Psi(-\alpha)} \, dF_{XX}(\alpha) \qquad (65)$$

and

$$X(\phi) = \int \Phi(-\alpha) \, dZ_X(\alpha) \qquad (66)$$

where c_X is an r vector, $c_{XX}(\cdot)$ is an $r \times r$ matrix of tempered distributions, $F_{XX}(\lambda)$ is a nonnegative matrix-valued measure satisfying

$$\int (1 + |\alpha|)^{-k} \, dF_{XX}(\alpha) < \infty \qquad (67)$$

for some nonnegative integer k, and finally $Z_X(\lambda)$ is a random function satisfying

$$E\, dZ_X(\lambda) = \delta(\lambda) \, c_X \, d\lambda \qquad (68)$$

$$\text{cov } \{dZ_X(\lambda), dZ_X(\mu)\} = \delta(\lambda - \mu) \, dF_{XX}(\lambda) \, d\mu. \qquad (69)$$

The spatial series of Section IX is a random Schwartz distribution corresponding to the functional

$$X(\phi) = \int X(t) \phi(t) \, dt$$

for $\phi \in \mathfrak{D}$. The representations indicated in that section may be deduced from the results of Theorem 1. It may be shown that k of (67) may be taken to be 0 for this case.

The stationary point process of Section XII is likewise a random Schwartz distribution corresponding to the functional

$$X(\phi) = \int \phi(t) \, dX(t)$$

for $\phi \in \mathfrak{D}$. The representations of Section XII may be deduced from Theorem 1. It may be shown that k of (67) may be taken to be 2 for this case.

Gelfand and Vilenkin [84] is a general reference to the theory of random Schwartz distributions. Theorem 1 is proved there.

A linear model that extends those of (42) and (58) to the present situation is one in which the $(r + s)$ vector-valued stationary random Schwartz distribution

$$\begin{bmatrix} S(\phi) \\ X(\phi) \end{bmatrix}$$

satisfies

$$E\, \{X(\phi) \,|\, S(\psi), \psi \in \mathfrak{D}\} = \mu \int \phi(t) \, dt + S(\phi * a)$$

$$= \mu \Phi(0) + \int \Phi(-\alpha) A(\alpha) \, dZ_S(\alpha). \quad (70)$$

In the case that the spectral measure is differentiable this last implies that

$$f_{XS}(\lambda) = A(\lambda) f_{SS}(\lambda) \qquad (71)$$

suggesting that the system may be identified if the spectral density may be estimated. We next set down a mixing assump-

tion, before constructing such an estimate and determining its asymptotic properties.

Given k variates $X_1, \cdots, X_k$ let cum $\{X_1, \cdots, X_k\}$ denote their joint cumulant or semi-invariant. Cumulants are defined and discussed in Kendall and Stuart [85] and Brillinger [20]. They are the elementary functions of the moments of the variates that vanish when the variates are independent. As such they provide measures of the degree of dependence of variates. We will make use of

Assumption 1. $X(\phi)$ is a stationary random Schwartz distribution with the property that for $\phi_1, \cdots, \phi_k \in \mathcal{S}$ and $a_1, \cdots, a_k = 1, \cdots, r; k = 2, 3, \cdots,$

$$\text{cum }\{X_{a_1}(\phi_1), \cdots, X_{a_k}(\phi_k)\} = \int \cdots \int \Phi_1(-\alpha^1) \cdots$$
$$\cdot \Phi_{k-1}(-\alpha^{k-1}) \Phi_k(\alpha^1 + \cdots + \alpha^{k-1})$$
$$\cdot f_{a_1 \cdots a_k}(\alpha^1, \cdots, \alpha^{k-1}) d\alpha^1 \cdots d\alpha^{k-1} \quad (72)$$

with

$$(1 + |\alpha^1|)^{-m_1} \cdots (1 + |\alpha^{k-1}|)^{-m_{k-1}} | f_{a_1 \cdots a_k}$$
$$\cdot (\alpha^1, \cdots, \alpha^{k-1}) | < L_k.$$

for some finite $m_1, \cdots, m_{k-1}, L_k$.

In the case that the spectral measure $F_{XX}(\lambda)$ is differentiable, relation (65) corresponds to the case $k = 2$ of (72). The character of Assumption 1 is one of limiting the size of the cumulants of the functionals of the process $X(\phi)$. It will be shown that it is a form of weak dependence requirement, for functionals of the process that are far apart in t, in the Appendix. The function $f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})$ appearing in (72) is called a *cumulant spectrum* of order k, see Brillinger [86] and the references therein. From (66) we see that it is also given by

$$\text{cum }\{dZ_{a_1}(\lambda^1), \cdots, dZ_{a_k}(\lambda^k)\} = \delta(\lambda^1 + \cdots + \lambda^k)$$
$$\cdot f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1}) d\lambda^1 \cdots d\lambda^k. \quad (73)$$

The fact that it only depends on $k - 1$ arguments results from the assumed stationarity of the process.

Let $\phi^{(T)}(t) = \phi(t/T)$ with $\phi \in \mathcal{D}$. As an analog of the Fourier transforms of Sections IX and XII we now define

$$d_X^{(T)}(\lambda) = X(\exp\{-i\langle\lambda, \cdot\rangle\} \phi^{(T)}) \quad (74)$$

for the stationary random Schwartz distribution $X(\phi)$. We can now state the following theorem.

Theorem 2: If Assumption 1 is satisfied, if $d_X^{(T)}(\lambda)$ is given by (74) and if $T | \lambda^j(T) - \lambda^k(T) | \to \infty$, $1 \leqslant j < k \leqslant J$, then 1)-4) of Section IX hold.

This theorem is proved in the Appendix. It provides a justification for the estimation procedures suggested in the paper and for the large sample approximations suggested for the distributions of the estimates.

We end this section by mentioning that a point process with events at positions τ_k, $k = 1, \cdots$ may be represented by the generalized function

$$\sum_k \delta(t - \tau_k)$$

the sampled function of Section III may be represented by the generalized function

$$\sum_{j=-\infty}^{\infty} X(j)\delta(t - j)$$

and that a point process with associated variate S may be represented by

$$\sum_k S_k \delta(t - \tau_k)$$

see Beutler and Leneman [87]. Mathéron [92] discusses the use of random Schwartz distributions in the smoothing of maps.

XVI. HIGHER ORDER SPECTRA AND NONLINEAR SYSTEMS

In the previous section we have introduced the higher order cumulant spectra of stationary random Schwartz distributions. In this section we will briefly discuss the use of such spectra and how they may be estimated.

In the case that the process under consideration is Gaussian, the cumulant spectra of order greater than two are identically 0. In the non-Gaussian case, the higher order spectra provide us with important information concerning the distribution of the process. For example were the process real-valued Poisson on the line with intensity c_N, then the cumulant spectrum of order k would be constant equal to $c_N(2\pi)^{1-k}$. Were the process the result of passing a series of independent identically distributed variates through a filter with transfer function $A(\lambda)$, then the cumulant spectrum of order k would be proportional to

$$A(\lambda^1) \cdots A(\lambda^{k-1}) A(-\lambda^1 - \cdots -\lambda^{k-1}).$$

Such hypotheses might be checked by estimating higher cumulant spectra.

An important use of higher order spectra is in the identification of polynomial systems such as those discussed in Wiener [88] and Brillinger [86] and Halme [89]. Tick [90] shows that if $S(t)$ is a stationary real-valued Gaussian series, if $\epsilon(t)$ is an independent stationary series and if the series $X(t)$ is given by

$$X(t) = \mu + \int a(t - u) S(u) \, du$$
$$+ \iint b(t - u, t - v) S(u) S(v) \, du \, dv + \epsilon(t) \quad (75)$$

then

$$f_{SX}(\lambda) = A(-\lambda) f_{SS}(\lambda)$$
$$f_{SSX}(\lambda, \mu) = 2B(-\lambda, -\mu) f_{SS}(\lambda) f_{SS}(\mu)$$

where

$$A(\lambda) = \int a(u) \exp\{-i\lambda u\} \, du$$

$$B(\lambda, \mu) = \iint b(u, v) \exp\{-i(\lambda u + \mu v)\} \, du \, dv$$

and $f_{SSX}(\lambda, \mu)$ is a third-order cumulant spectrum. It follows that both the linear transfer function $A(\lambda)$ and the bitransfer function $B(\lambda, \mu)$ of the system may be estimated, from estimates of second- and third-order spectra, following the probing of the system by a single Gaussian series. References to the identification of systems of order greater than 2, and to the case of non-Gaussian $S(t)$ are given in [86].

We turn to the problem of constructing an estimate of a kth order cumulant spectrum. In the course of the proof of

207

Theorem 2 given in the Appendix, we will see that

$$\text{cum } \{d_{a_1}^{(T)}(\lambda^1), \cdots, d_{a_k}^{(T)}(\lambda^k)\} \sim \begin{cases} (2\pi)^{p(k-1)} \int \phi^{(T)}(t)^k \, dt \, f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1}), & \text{if } \lambda^1 + \cdots + \lambda^{k-1} = 0 \\ 0, & \text{if } \lambda^1 + \cdots + \lambda^{k-1} \neq 0. \end{cases}$$

Suppose that no proper subset of $\lambda^1, \cdots, \lambda^k$ sums to 0. It then follows from the principal relation connecting moments and cumulants that

$$E\{d_{a_1}^{(T)}(\lambda^1) \cdots d_{a_k}^{(T)}(\lambda^k)\} \sim (2\pi)^{p(k-1)} \int \phi^{(T)}(t)^k$$
$$\cdot dt \, f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})$$

provided $\lambda^1 + \cdots + \lambda^k = 0$. This last one suggests the use of kth order periodogram

$$I_{a_1 \cdots a_k}^{(T)}(\lambda^1, \cdots, \lambda^{k-1}) = (2\pi)^{-p(k-1)} \left(\int \phi^{(T)}(t)^k dt \right)^{-1}$$
$$\cdot d_{a_1}^{(T)}(\lambda^1) \cdots d_{a_{k-1}}^{(T)}(\lambda^{k-1})$$
$$\times d_{a_k}^{(T)}(-\lambda^1 - \cdots - \lambda^{k-1}) \qquad (76)$$

as a naive estimate of the spectrum $f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})$ provided that no proper subset of $\lambda^1, \cdots, \lambda^{k-1}$ sums to 0. From what we have seen in the case $k = 2$ this estimate will be unstable. It follows that we should in fact construct an estimate by smoothing the periodogram (76) over $(k - 1)$-tuples of frequencies in the neighborhood of $\lambda^1, \cdots, \lambda^{k-1}$, but such that no proper subset of the $(k - 1)$-tuple sums to 0. Details of this construction are given in Brillinger and Rosenblatt [91] for the discrete time case. We could equally well have constructed an estimate using the Fourier transforms $d_X^{(T)}(\lambda, j)$ based on disjoint domains.

APPENDIX

We begin by providing a motivation for Assumption 1 of Section XIV. Suppose that

$$\text{cum } \{X_{a_1}(\phi_1), \cdots, X_{a_k}(\phi_k)\}, \qquad \phi_1, \cdots, \phi_k \in \mathfrak{D}$$

is continuous in each of its arguments. Being a continuous multilinear functional it can be written

$$c_{a_1 \cdots a_k}(\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_k)$$

where $c_{a_1 \cdots a_k}$ is a Schwartz distribution on $\mathfrak{D}(R^{pk})$, from the Schwartz nuclear theorem. If the process is stationary this distribution satisfies

$$c_{a_1 \cdots a_k}(S^u \phi_1 \otimes S^u \phi_2 \otimes \cdots \otimes S^u \phi_k) = c_{a_1 \cdots a_k}$$
$$\cdot (\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_k).$$

It follows that it has the form

$$\mathcal{C} \left(\int \phi(t + u^1, \cdots, t + u^{k-1}, t) \, dt \right)$$

for $\phi \in \mathfrak{D}(R^{pk})$ where $\mathcal{C}$ is a distribution on $\mathfrak{D}(R^{p(k-1)})$.

Now consider the case in which the process $X(\phi)$ has the property that

$$\text{cum } \{X_{a_1}(\phi_1), \cdots, X_{a_k}(\phi_k)\} = 0$$

when the supports of $\phi_1, \cdots, \phi_{k-1}$ are farther away from that of ϕ_k than some number ρ. This means that the distribution $\mathcal{C}$ has compact support. By the Schwartz–Paley–Wiener theorem, $\mathcal{C}$ is, therefore, the Fourier transform of a function of slow growth, say $f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})$ and we may write the relation (72). In the case that values of the process $X(\phi)$ at a distance from each other are only weakly dependent, we can expect the cumulant to be small and for the representation (72) to hold with (73) satisfied.

Proof of Theorem 2: We see from (66) and (73)

$$\text{cum } \{d_{a_1}^{(T)}(\lambda^1), \cdots, d_{a_k}^{(T)}(\lambda^k)\}$$

$$= \int \cdots \int \Phi_1^{(T)}(\alpha^1 - \lambda^1) \cdots \Phi_{k-1}^{(T)}(\alpha^{k-1} - \lambda^{k-1}) \Phi_k^{(T)}$$
$$\cdot (-\alpha^1 - \cdots - \alpha^{k-1} - \lambda^k) f_{a_1 \cdots a_k}(\alpha^1, \cdots, \alpha^{k-1})$$
$$\cdot d\alpha^1 \cdots d\alpha^{k-1}$$

$$= T^p \int \cdots \int \Phi_1(\beta^1) \cdots \Phi_{k-1}(\beta^{k-1}) \Phi_k$$
$$\cdot (-\beta^1 - \cdots - \beta^{k-1} - T(\lambda^1 + \cdots + \lambda^k) f_{a_1 \cdots a_k}$$
$$\cdot (\lambda^1 + T^{-1}\beta^1, \cdots, \lambda^{k-1} + T^{-1}\beta^{k-1}) d\beta^1 \cdots d\beta^{k-1}$$

$$\sim T^p \int \cdots \int \Phi_1(\beta^1) \cdots \Phi_{k-1}(\beta^{k-1}) \Phi_k(-\beta^1 - \cdots - \beta^{k-1})$$
$$\cdot d\beta^1 \cdots d\beta^{k-1} f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1}),$$
$$\text{for } \lambda^1 + \cdots + \lambda^k = 0$$

$$= 0(T^p), \qquad \text{for } \lambda^1 + \cdots + \lambda^k \neq 0.$$

It follows from this last that the standardized joint cumulants of order greater than 2 tend to 0 and so the Fourier transforms are asymptotically normal.

REFERENCES

[1] G. G. Stokes, "Note on searching for periodicities," *Proc. Roy. Soc.*, vol. 29, p. 122, 1879.
[2] A. Schuster, "The periodogram of magnetic declination," *Cambridge Phil. Soc.*, vol. 18, p. 18, 1899.
[3] E. T. Whittaker and A. Robinson, *The Calculus of Observations.* Cambridge, England: Cambridge Univ. Press, 1944.
[4] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, 1965.
[5] A. A. Michelson, *Light Waves and Their Uses.* Chicago, Ill.: Univ. Chicago Press, 1907.
[6] I. Newton, *Opticks.* London, England: W. Innys, 1730.
[7] M. I. Pupin, "Resonance analysis of alternating and polyphase currents," *Trans. AIEE*, vol. 9, p. 523, 1894.
[8] H. L. Moore, *Economic Cycles Their Law and Cause.* New York: Macmillan, 1914.
[9] H. A. Panofsky and R. A. McCormick, "Properties of spectra of atmospheric turbulence at 100 metres," *Quart. J. Roy. Meteorol. Soc.*, vol. 80, pp. 546–564, 1954.
[10] J. A. Leese and E. S. Epstein, "Application of two-dimensional spectral analysis to the quantification of satelite cloud photographs," *J. Appl. Meteorol.*, vol. 2, pp. 629–644, 1963.
[11] M. S. Bartlett, "The spectral analysis of point processes," *J. Roy. Stat. Soc.*, vol. B 25, pp. 264–296, 1963.

[12] —, "The spectral analysis of two dimensional point processes," *Biometrika*, vol. 51, pp. 299–311, 1964.

[13] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, pp. 1264–1291, 1968.

[14] B. P. Bogert, M. J. Healey, and J. W. Tukey, "The quefrency alanysis of time series for echoes: cepstrum, pseudo-covariance, cross-cepstrum and saphe cracking," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 209–243, 1963.

[15] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, pp. 1–16, 1950.

[16] R. H. Jones, "A reappraisal of the periodogram in spectral analysis," *Technometrics*, vol. 7, pp. 531–542, 1965.

[17] K. Hasselman, W. Munk, and G. J. F. MacDonald, "Bispectra of ocean waves," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 125–139, 1963.

[18] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258, 1930.

[19] H. Wold, *Bibliography on Time Series and Stochastic Processes*. London, England: Oliver and Boyd, 1965.

[20] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston, 1974.

[21] A. Ya. Khintchine, "Korrelationstheories der stationären stochastischen Prozesse," *Math. Ann.*, vol. 109, pp. 604–615, 1934.

[22] H. Cramér, "On harmonic analysis in certain functional spaces," *Ark. Mat. Astron. Fys.*, vol. 28B, pp. 1–7, 1942.

[23] V. P. Leonov and A. N. Shiryaev, "Some problems in the spectral theory of higher moments, II," *Theory Prob. Appl. (USSR)*, vol. 5, pp. 460–464, 1960.

[24] B. Picinbono, "Tendence vers le caractère Gaussien par filtrage seléctif," *C. R. Acad. Sci. Paris*, vol. 248, p. 2280, 1959.

[25] M. Rosenblatt, "Some comments on narrow band-pass filters," *Quart. Appl. Math.*, vol. 18, pp. 387–393, 1961.

[26] D. R. Brillinger, "The frequency analysis of relations between stationary spatial series," in *Proc. 12th Biennial Seminar of the Canadian Math. Congress*, R. Pyke, Ed. Montreal, P.Q., Canada: Can. Math. Congr., pp. 39–81, 1970.

[27] E. J. Hannan and P. J. Thomson, "Spectral inference over narrow bands," *J. Appl. Prob.*, vol. 8, pp. 157–169, 1971.

[28] E. Parzen, "On consistent estimates of the spectrum of stationary time series," *Ann. Math. Statist.*, vol. 28, pp. 329–348, 1957.

[29] M. Rosenblatt, "Statistical analysis of stochastic processes with stationary residuals," in *Probability and Statistics*, U. Grenander, Ed. New York: Wiley, pp. 246–275, 1959.

[30] W. D. Wright, *The Measurement of Color*. New York: Macmillan. 1958.

[31] E. W. Carpenter, "Explosions seismology," *Science*, vol. 147, pp. 363–373, 1967.

[32] D. G. Lambert, E. A. Flinn, and C. B. Archambeau, "A comparative study of the elastic wave radiation from earthquakes and underground explosions," *Geophys. J. Roy. Astron. Soc.*, vol. 29, pp. 403–432, 1972.

[33] W. H. Munk and G. J. F. MacDonald, *Rotation of the Earth*. Cambridge, England: Cambridge Univ. Press, 1960.

[34] G. J. F. MacDonald and N. Ness, "A study of the free oscillations of the Earth," *J. Geophys. Res.*, vol. 66, pp. 1865–1911, 1961.

[35] R. L. Wegel and C. R. Moore, "An electrical frequency analyzer," *Bell Syst. Tech. J.*, vol. 3, pp. 299–323, 1924.

[36] N. Wiener, *Time Series*. Cambridge, Mass.: M.I.T. Press, 1964.

[37] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.

[38] E. Parzen, "An approach to empirical time series analysis," *Radio Sci.*, vol. 68 D, pp. 551–565, 1964.

[39] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, 1971.

[40] J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375–376, 1972.

[41] K. N. Berk, "Consistent autoregressive spectral estimates," *Ann. Stat.*, vol. 2, pp. 489–502, 1974.

[42] V. E. Pisarenko, "On the estimation of spectra by means of nonlinear functions of the covariance matrix," *Geophys. J. Roy. Astron. Soc.*, vol. 28, pp. 511–531, 1972.

[43] J. Capon, "Investigation of long-period noise at the large aperture seismic array," *J. Geophys. Res.*, vol. 74, pp. 3182–3194, 1969.

[44] E. J. Hannan, *Multiple Time Series*. New York: Wiley, 1970.

[45] T. W. Anderson, *The Statistical Analysis of Time Series*. New York: Wiley, 1971.

[46] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat. Astron. Fys.*, vol. 2, pp. 423–434, 1953.

[47] J. M. Chambers, "Fitting nonlinear models: numerical techniques," *Biometrika*, vol. 60, pp. 1–14, 1973.

[48] D. R. Brillinger, "An empirical investigation of the Chandler wobble and two proposed excitation processes," *Bull. Int. Stat. Inst.*, vol. 39, pp. 413–434, 1973.

[49] P. Whittle, "Gaussian estimation in stationary time series," *Bull. Int. Stat. Inst.*, vol. 33, pp. 105–130, 1961.

[50] A. M. Walker, "Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary nondeterministic time-series," *J. Australian Math. Soc.*, vol. 4, pp. 363–384, 1964.

[51] K. O. Dzaparidze, "A new method in estimating spectrum parameters of a stationary regular time series," *Teor. Veroyat. Ee Primen.*, vol. 19, p. 130, 1974.

[52] L. H. Koopmans, "On the coefficient of coherence for weakly stationary stochastic processes," *Ann. Math. Stat.*, vol. 35, pp. 532–549, 1964.

[53] H. Akaike and Y. Yamanouchi, "On the statistical estimation of frequency response function," *Ann. Inst. Stat. Math.*, vol. 14, pp. 23–56, 1962.

[54] L. J. Tick, "Estimation of coherency," in *Advanced Seminar on Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 133–152.

[55] R. J. Bhansali, "Estimation of the Wiener filter," in *Contributed Papers 39th Session Int. Stat. Inst.*, vol. 1, pp. 82–88, 1973.

[56] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.

[57] B. R. Hunt, "Biased estimation for nonparametric identification of linear systems," *Math. Biosci.*, vol. 10, pp. 215–237, 1971.

[58] S. Bochner and K. Chandrasekharan, *Fourier Transforms*. Princeton, N.J.: Princeton Univ. Press, 1949.

[59] J. Mannos, "A class of fidelity criteria for the encoding of visual images," Ph.D. dissertation, Univ. California, Berkeley, 1972.

[60] L. J. Tick, "Conditional spectra, linear systems and coherency," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 197–203, 1963.

[61] C. W. J. Granger, *Spectral Analysis of Economic Time Series*. Princeton, N.J.: Princeton Univ. Press, 1964.

[62] N. R. Goodman, "Measurement of matrix frequency response functions and multiple coherence functions," Air Force Dynamics Lab., Wright Patterson AFB, Ohio, Tech. Rep. AFFDL-TR-65-56, 1965.

[63] J. S. Bendat and A. Piersol, *Measurement and Analysis of Random Data*. New York: Wiley, 1966.

[64] G. W. Groves and E. J. Hannan, "Time series regression of sea level on weather," *Rev. Geophys.*, vol. 6, pp. 129–174, 1968.

[65] W. Gersch, "Causality or driving in electrophysiological signal analysis," *J. Math. Biosci.*, vol. 14, pp. 177–196, 1972.

[66] N. R. Goodman, "Eigenvalues and eigenvectors of spectral density matrices," Tech. Rep. 179, Seismic Data Lab., Teledyne, Inc., 1967.

[67] D. R. Brillinger, "The canonical analysis of time series," in *Multivariate Analysis–II*, P. R. Krishnaiah, Ed. New York: Academic, pp. 331–350, 1970.

[68] M. B. Priestley, T. Subba Rao, and H. Tong, "Identification of the structure of multivariable stochastic systems," in *Multivariate Analysis–III*, P. R. Krishnaiah, Ed. New York: Academic, pp. 351–368, 1973.

[69] M. Miyata, "Complex generalization of canonical correlation and its application to a sea-level study," *J. Marine Res.*, vol. 28, pp. 202–214, 1970.

[70] W. S. Ligett, Jr., "Passive sonar: Fitting models to multiple time series," paper presented at NATO Advanced Study Institute on Signal Processing, Loughborough, U. K., 1972.

[71] D. R. Brillinger, "The analysis of time series collected in an experimental design," *Multivariate Analysis–III*, P. R. Krishnaiah, Ed. New York: Academic, pp. 241–256, 1973.

[72] D. R. Brillinger and M. Hatanaka, "An harmonic analysis of nonstationary multivariate economic processes," *Econometrica*, vol. 35, pp. 131–141, 1969.

[73] E. J. Hannan and R. D. Terrell, "Multiple equation systems with stationary errors," *Econometrica*, vol. 41, pp. 299–320, 1973.

[74] C. R. Rao, *Linear Statistical Inference and its Applications*. New York: Wiley, 1965.

[75] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*. London, England: Methuen, 1966.

[76] D. R. Brillinger, "The spectral analysis of stationary interval functions," *Proc. 6th Berkeley Symp. Math. Stat. Prob. Vol. 1*, L. M. Le Cam, J. Neyman, and E. L. Scott, Eds. Berkeley, Calif.: Univ. California Press, pp. 483–513, 1972.

[77] D. J. Daley and D. Vere-Jones, "A summary of the theory of point processes," in *Stochastic Point Processes*, P. A. W. Lewis, Ed. New York: Wiley, pp. 299–383, 1972.

[78] L. Fisher, "A survey of the mathematical theory of multidimensional point processes," in *Stochastic Point Processes*, P. A. W. Lewis, Ed. New York: Wiley, pp. 468–513, 1972.

[79] A. G. Hawkes and L. Adamopoulos, "Cluster models for earthquakes—regional comparisons," *Bull. Int. Stat. Inst.*, vol. 39, pp. 454–460, 1973.

[80] A. M. Yaglom, "Some classes of random fields in n-dimensional space related to stationary random processes," *Theory Prob. Appl. (USSR)*, vol. 2, pp. 273–322, 1959.

[81] L. Schwartz, *Théorie des Distributions, Vols. 1, 2*. Paris, France: Hermann, 1957.

Solid-State Control of Electric Drives

ROBERT G. SCHIEMAN, MEMBER, IEEE, EDWARD A. WILKES, AND
HOWARD E. JORDAN, SENIOR MEMBER, IEEE

Invited Paper

Abstract—A tutorial review of the dc and ac electric-drive field is presented. The goal is to present fundamental concepts, principle concerns, and key developments in electric-drive technology. Principles of ac and dc power converters and ac and dc motors are presented. Then the combination of the converter and motor to provide a complete drive system is discussed along with drive-system characteristics and methods for analyzing performance. Finally, some application guidelines for both ac and dc systems are given.

I. Introduction

THE GROWTH of electric drives has closely paralleled the growth of automation in industry. Electric-drive systems provide a convenient means for controlling the operation of industrial machinery. The high reliability and great versatility of electric drives has resulted in their widespread application. In size, electric drives range all the way from fractions of one horsepower up to thousands of horsepower. Speeds range from stalled positioning systems up to 15 000 rev/min and higher.

Historically, the first electric-drive system to gain real prominence was the Ward Leonard System, patented by H. Ward Leonard in the 1890's. The history of dc electric drives proceeded from the basic Ward Leonard principle to various modifications thereof, in approximately the following steps:

1) rheostat control of generator field;
2) tandem field rheostat control of generator field and motor field;

3) thyratron control of generator and motor fields and later thyraton control of the armature voltage of small dc motors;
4) ignitron and mercury pool control of the armature voltage of dc machines too large for thyratrons;
5) magnetic amplifier control of generator field and motor armature voltage; and
6) thyristor control of generator and motor fields and later thyristor control of armature voltage.

During the latter part of the era of the thyratron, the transistor started to replace vacuum tubes in drive regulators. Now solid-state electronic circuits are used to implement special compensating circuits that significantly improve feedback control system response. Microelectronic circuits, particularly operational amplifiers, are used extensively in drive systems today. The operational amplifier circuits are the key to drive-system response, stability, and regulation.

The ac motor variable speed drive development is very similar to the dc.

Initially, the motor alternator set with field rheostats was used to control the ac motor speed. Then other methods of ac motor control were developed. They are as follows:

1) wound rotor resistance control to vary speed with torque load;
2) methods of replacing the resistor in the rotor with other rotating machinery or rectifiers to pump the power back into the ac line;
3) ac motor stator voltage control by the use of resistors, reactors, magnetic amplifiers, thyratrons, ignitrons, mercury-pool tubes or thyristors; and
4) replacement of the motor-alternator set for varying voltage and frequency to the motor with static devices.